On Central Difference Approximations to General Second Order Elliptic Equations*

W. Layton†
Department of Mathematics and Statistics
University of Pittsburgh
Pittsburgh, Pennsylvania 15260

and

T. D. Morley
School of Mathematics
Georgia Institute of Technology
Atlanta, Georgia 30332

Submitted by Richard S. Varga

ABSTRACT

Consider the second order elliptic equation

$$Lu := -au_{xx} - 2bu_{xy} - cu_{yy} + d(x,y)u = f(x,y) \tag{1}$$

in $[0,1]\times[0,1]$, with periodic boundary conditions, and

$$b^2 < ac, \qquad a > 0, \qquad c > 0, \qquad d(x,y) \geq 0.$$

Finite difference discretizations require a much stronger condition than ellipticity to give a scheme of positive type. In this paper, it is shown that the standard central difference discretization of (1) is of *monotone type* although it is not positive type. Specifically, the inverse matrix arising from it has one sign.

INTRODUCTION

In this paper we prove that the standard central difference discretization of the second order elliptic equation (subject to periodic boundary conditions

*The work of the first author was partially supported by AFOSR grant number 83-0101 while he was visiting the Mathematics Department of Carnegie Mellon University.

†On leave from School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332.

 0024-3795/87/$3.50

on u and $\partial u/\partial n$)

$$Lu := -au_{xx} - 2bu_{xy} - cu_{yy} + d(x,y)u = f(x,y) \quad \text{in} \quad \Omega = [0,1]\times[0,1], \quad (1)$$

$$u(0,y) = u(1,y), \qquad u(x,0) = u(x,1), \qquad 0 \leqslant x, y \leqslant 1,$$

$$u_x(0,y) = u_x(1,y), \qquad u_y(x,0) = u_y(x,1), \qquad 0 \leqslant x, y \leqslant 1,$$

is (inverse) monotone. If A is the discretization matrix associated with L via some node ordering, we show that $A^{-1} \geqslant 0$ elementwise. In (1) the constants a, b, c satisfy the ellipticity condition, and $d(x,y)$ satisfies a weak condition to ensure invertibility of L and A:

$$b^2 < ac, \qquad a > 0, \qquad c > 0,$$

$$d(x,y) \geqslant 0 \qquad \text{in } \Omega,$$

$$d(x,y) > 0 \qquad \text{in some subregion of } \Omega \text{ containing at least one mesh point.} \quad (2)$$

This result is interesting and useful, since the standard (9-point) central difference discretization of (1.1) is never of positive type when $b \neq 0$. Other common $O(h^2)$ discretizations of (1) are of positive type (Mitchell and Griffiths [11]) provided the much more restrictive condition holds:

$$|b| < \min\{|a|, |c|\}. \quad (3)$$

In [2] Bramble and Hubbard show that in the interior submatrix of the standard $O(h^2)$ discretization of

$$u_{xx} + u_{xy} + u_{yy} = f \qquad \text{in } \Omega,$$

$$u = 0 \qquad \text{on } \partial\Omega,$$

has a positive inverse (once the boundary unknowns are eliminated from the linear system). A careful analysis of their proof shows that it extends to the more general elliptic problem

$$au_{xx} + 2bu_{xy} + cu_{yy} = f \qquad \text{in } \Omega,$$

$$u = 0 \qquad \text{on } \partial\Omega; \qquad a, b, c \geqslant 0,$$

where a, b, and c are constants, under the condition that

$$b \leqslant \frac{ac}{a+c}.$$

Note that the above condition is implied by the condition

$$b \leqslant \tfrac{1}{2}\min(a, c).$$

It is easy to adapt the techniques of [2] to the periodic case. Thus the following is essentially contained in [2]:

THEOREM. *The standard $O(h^2)$ central difference approximation L^h, given below, to* (1) *is invertible and satisfies $A^{-1} \geqslant 0$ elementwise if $b \leqslant ac/(a+c)$.*

By the harmonic-geometric mean inequality we have (for $a, b, c \geqslant 0$)

$$\frac{ac}{a+c} < 2\frac{ac}{a+c} \leqslant \sqrt{ac},$$

and thus the above condition is strictly stronger than ellipticity. For example, the above theorem does *not* apply to the problem

$$u_{xx} + 2bu_{xy} + 2u_{yy} = f$$

if $b > \frac{2}{3}$. Here ellipticity follows from the weaker requirement that $|b| < \sqrt{2}$. Further, one can ask if there exists *any* 9-point approximation to (1) of the form

$$L^h u_{ij} \equiv \sum_{\alpha,\beta \in \{-1,0,+1\}} d_{\alpha,\beta} T_{\alpha,\beta} u_{ij} = f_{ij}$$

that is consistent and of positive type provided only $b^2 < ac$. On a uniform mesh the answer is no; Greenspan and Jain [10] have shown that (3) is also necessary in the case.

Let $h = 1/N$, $x_j = jh$, $y_k = kh$, $j, k = 0, \ldots, N$. We identify x_0 with x_N, and y_0 with y_N. Define, as usual, the difference operators

$$D_x^+ u_{ij} = \frac{u_{i+1,j} - u_{ij}}{h}, \qquad D_y^+ u_{ij} = \frac{u_{ij+1} - u_{ij}}{h},$$

$$D_x^- u_{ij} = \frac{u_{ij} - u_{i-1,j}}{h}, \qquad D_y^- u_{ij} = \frac{u_{ij} - u_{i,j-1}}{h}.$$

From these, we obtain the usual $O(h^2)$ accurate approximation to second derivatives:

$$D_x^2 = D_x^+ D_x^-, \qquad D_y^2 = D_y^+ D_y^-,$$

$$D_{xy}^+ = \tfrac{1}{2}\left(D_x^+ D_y^+ + D_x^- D_y^-\right),$$

$$D_{xy}^- = \tfrac{1}{2}\left(D_x^+ D_y^- + D_x^- D_y^+\right),$$

$$D_{xy}^\theta = \left(\theta D_{xy}^+ + (1-\theta) D_{xy}^-\right), \qquad 0 \leqslant \theta \leqslant 1,$$

$$D_{xy} = D_{xy}^{(1/2)}$$

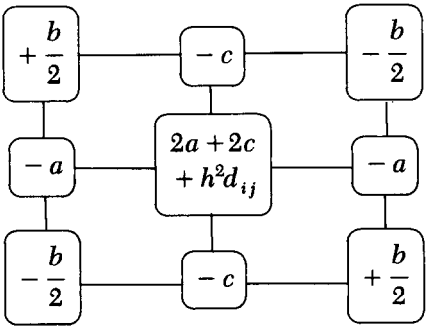
We define the translation operators $T_{\alpha\beta}$ as usual also:

$$T_{\alpha\beta} u(x, y) = u(x + \alpha h, y + \beta h).$$

The approximation scheme for $Lu = f$ is the 9-point central difference approximation:

$$L_\theta^h u_{ij} := -aD_x^2 u_{ij} - 2bD_{xy}^\theta u_{ij} - cD_y^2 u_{ij} + d_{ij} u_{ij} = f_{ij}.$$

Henceforth, we let L^h denote the usual difference approximation to L, $L^h := L_\theta^h|_{\theta = 1/2}$, and we let A denote the associated discretization matrix which is attached to L^h via some global node ordering of the $(x_j, y_k) \in [0,1] \times [0,1]$. Note that A is a circulant matrix although it is *not* an M-matrix, since the required sign pattern is violated. $h^2 * L^h$ is represented by the following difference molecule:



The theorem that is proven is:

THEOREM 1. *Suppose* (2) *holds. Then* A *is invertible and* $A^{-1} \geqslant 0$ *elementwise, so* L^h *is an (inverse) monotone scheme. The same result holds for* L_θ^h *provided* $0 \leqslant \theta \leqslant 1$.

The techniques used for proving the above are based upon the early work of Varga [13] on regular splittings and upon a "positive type decomposition" deduced by Brandt [4]. The connection between the theory of regular splittings and the work of Bramble and Hubbard [2, 3] on monotone difference schemes was exploited by Price [12] in a similar manner to ours.

2. PROOF OF THEOREM 1

The difference scheme L^h has a sign pattern that can arise from the composition of two positive type methods. Exploiting this, we will show that A factors into the product of two M-matrices modulo a diagonal error term which has a specified sign. All matrix and vector inequalities are to be interpreted elementwise.

DEFINITION. An $n \times n$ matrix A is called

(i) a *monotone matrix* if A^{-1} exists and $Ax \leqslant Ay$ implies $x \leqslant y$; equivalently, if $A^{-1} \geqslant 0$;

(ii) an *L-matrix* if $A = (a_{ij})$ with $a_{ii} \geqslant 0$ and $a_{ij} \leqslant 0$ for $i \neq j$

(iii) an *M-matrix* if A is both an L-matrix and a monotone matrix.

We now summarize a few properties of M-matrices that we will be using. For proofs and more detail, see Varga [13].

PROPOSITION 1. *Suppose that* A *is an L-matrix.*

(a) *If* A *is strictly diagonally dominant, then* A *is an M-matrix.*

(b) *If* A *is strictly diagonally dominant and irreducible, then* A *is an M-matrix and* $A^{-1} > 0$.

(c) *If* A *is irreducible, diagonally semidominant with at least one row of* A *strictly dominant — that is, for some* i

$$a_{ii} > \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|$$

— then A *is an M-matrix and, in fact,* $A^{-1} > 0$.

PROPOSITION 2. *Suppose* (2) *holds. Then there is a positive constant* λ^2 *of order* 1 *such that if* $h^2 d < \lambda^2$, A *factors as*

$$A = M_1 M_2 - R.$$

M_1, M_2 *are strictly diagonally dominant, irreducible M-matrices arising from the positive type difference operators*:

$$\Lambda_1 u_{ij} = \lambda_0 u_{ij} - \lambda_1 T_{10} u_{ij} - \lambda_2 T_{11} u_{ij} - \lambda_3 T_{01} u_{ij},$$

$$\Lambda_2 u_{ij} = \lambda_0 u_{ij} - \lambda_1 T_{-10} u_{ij} - \lambda_2 T_{-1-1} u_{ij} - \lambda_3 T_{0-1} u_{ij},$$

$$\lambda_j \geqslant 0, \qquad \lambda := \lambda_1 - \lambda_1 - \lambda_2 - \lambda_3 > 0,$$

respectively, under the same node ordering as A. *The* λ_j*'s are given below, and* $R > 0$ *arises from the operator*

$$\Delta u_{ij} = \lambda^2 u_{ij} - h^2 d_{ij} u_{ij}.$$

REMARK. This operator decomposition is given in §4 of Brandt [4]. We include a proof for completeness and because we believe this calculation to be elegant and perhaps useful in other contexts.

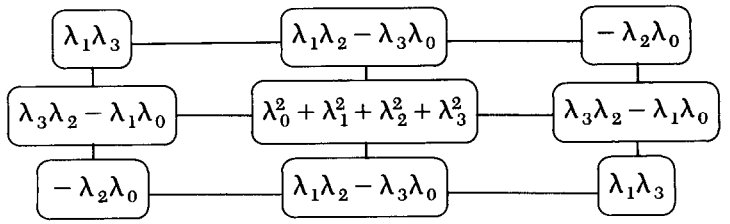
Proof. Suppose $b > 0$. Then at (x_i, y_j) define Λ_1, Λ_2 by:

$$\Lambda_1 u_{ij} = \lambda_0 u_{ij} - \lambda_1 T_{10} u_{ij} - \lambda_2 T_{11} u_{ij} - \lambda_3 T_{01} u_{ij},$$

$$\Lambda_2 u_{ij} = \lambda_0 u_{ij} - \lambda_1 T_{-10} u_{ij} - \lambda_2 T_{-1-1} u_{ij} - \lambda_3 T_{0-1} u_{ij},$$

where $\lambda_j \geqslant 0$, and define $\lambda = \lambda_0 - \lambda_1 - \lambda_2 - \lambda_3$.

It is straightforward to see $\Lambda_1 \Lambda_2$ gives the difference stencil



which has the same sign pattern as L^h. No difficulty arises in the composition, since we are now in the constant coefficient case on a domain without

boundary (the 2-torus). We determine the λ_j's by requiring that $\Lambda_1\Lambda_2$ agree with h^2L^h away from (x_i, y_j). This will then ensure that M_1M_2 agrees with A except for the diagonal. Comparing the two stencils gives the equations

$$\begin{aligned} \lambda_0\lambda_2 &= b/2, & \lambda_1\lambda_3 &= b/2, \\ \lambda_0\lambda_1 - \lambda_2\lambda_3 &= a, & \lambda_0\lambda_3 - \lambda_2\lambda_1 &= c. \end{aligned} \tag{4}$$

These equations are solved as follows. Multiplying the third equation by $\lambda_0\lambda_1$ and using the first two gives

$$(\lambda_0\lambda_1)^2 - a\lambda_0\lambda_1 - \frac{b^2}{4} = 0,$$

whence

$$\lambda_0\lambda_1 = \tfrac{1}{2}\left[(a^2 + b^2)^{1/2} + a\right] > 0,$$

and similarly,

$$\lambda_2\lambda_3 = \tfrac{1}{2}\left[(a^2 + b^2)^2 - a\right] > 0,$$

$$\lambda_0\lambda_3 = \tfrac{1}{2}\left[(c^2 + b^2)^{1/2} + c\right] > 0,$$

$$\lambda_2\lambda_1 = \tfrac{1}{2}\left[(c^2 + b^2)^{1/2} - c\right] > 0.$$

The λ_i are then determined through

$$\lambda_i^2 = \frac{\lambda_i\lambda_j \cdot \lambda_i\lambda_k}{\lambda_j\lambda_k}.$$

It is straightforward to check that, after some algebraic manipulation, these give unique and consistent solutions for $\lambda_{0,1,2,3}$ and that these λ_j satisfy the original equations.

We now show that each $M_{1,2}$ is strictly diagonally dominant, i.e., $\lambda = \lambda_0 - \lambda_1 - \lambda_2 - \lambda_3 > 0$. Indeed, from the definition of the λ_j's we have

$$a = (\lambda_0 - \lambda_3)(\lambda_2 + \lambda_1) \qquad (\text{thus} \quad \lambda_0 > \lambda_3),$$

$$b = (\lambda_0\lambda_2 + \lambda_1\lambda_3)$$

$$c = (\lambda_0 - \lambda_1)(\lambda_2 + \lambda_3),$$

from which we calculate

$$0 < ac - b^2 = \lambda(\lambda_0\lambda_1\lambda_2 + \lambda_0\lambda_1\lambda_3 + \lambda_0\lambda_2\lambda_3 - \lambda_1\lambda_2\lambda_3).$$

Since $\lambda_0 > \lambda_3$, $\lambda_0\lambda_1\lambda_2 - \lambda_1\lambda_2\lambda_3 > 0$ and hence $\lambda > 0$. Thus we have defined M_1 and M_2 at points where $b > 0$ such that M_1 and M_2 are strictly diagonally dominant M-matrices and $M_1M_2 - A$ is a diagonal matrix.

When $b < 0$, we define Λ_1 and Λ_2 via the difference operators

$$\Lambda_1 u_{ij} = \lambda_0 u_{ij} - \lambda_1 T_{+10} u_{ij} - \lambda_2 T_{1-1} u_{ij} - \lambda_3 T_{0-1} u_{ij},$$

$$\Lambda_2 u_{ij} = \lambda_0 u_{ij} - \lambda_1 T_{-10} u_{ij} - \lambda_2 T_{-11} u_{ij} - \lambda_3 T_{01} u_{ij},$$

and proceed, *mutatis mutandis*, as in the case $b > 0$.

Before we show that M_1 and M_2 are irreducible we note that they are so precisely because the boundary conditions are periodic.

Consider the case when $b > 0$. The directed graph associated with M_2 is given, at a representative node, by Figure 1. Thus, on, e.g., a 5×5 mesh we would have Figure 2 (omitting cycles for clarity). *The sides* $x = 0$ *and* $x = 1$ *are identified, as are* $y = 0$ *and* $y = 1$.

Since the graph of M_2 is on a 2-torus, we have wraparound connections. Further, all four corners of the square are identified. From any meshpoint we may follow the graph to (0,0) (see Figure 2), which is identified with (1,1), and from (1,1) to any other mesh point.

Thus M_2 is irreducible. A similar argument holds for M_1 by flipping Figure 2 appropriately. ■

We now complete the proof of the main result. For this, our strategy is to show that the spectral radius of $M_2^{-1}M_1^{-1}R$, is smaller than one: $\rho(M_2^{-1}M_1^{-1}R) < 1$. The theorem will then follow from the following result of Price [12, Theorem 2.2, p. 490] (with the identification $M = M_1M_2$).

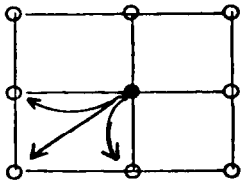


FIG. 1. The directed graph of one row of M_2.

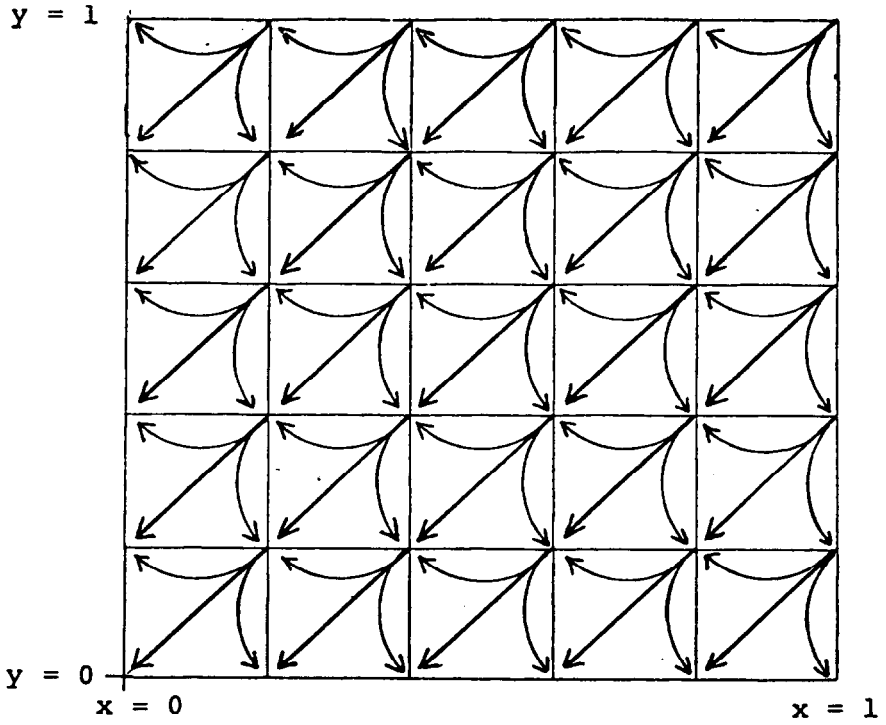


FIG. 2. The directed graph associated with M_2 or an $S \times S$ mesh.

THEOREM 2 (Price [12, Theorem 2.2]). *Let A be a real $n \times n$ matrix. Then A is monotone if and only if there exists a real $n \times n$ matrix R with the following three properties:*

(i) $M = A + R$ *is monotone.*
(ii) $R \geqslant 0$
(iii) $\rho(M^{-1}R) < 1$.

PROPOSITION 3. *Under the assumptions of Theorem* 1, *for* $h^2 d_{ij} < \lambda$, $\rho(M_2^{-1}M_1^{-1}R) < 1$.

Proof. Define the vector $e_j = 1$ for all j. Let (x_I, y_J) be such that $d_{IJ} > 0$. Define ξ by

$$\xi_{IJ} = 1, \qquad \xi_{kl} = 0 \quad \text{for all other } k, l.$$

We find, by direct calculation, that $A\mathbf{e} \geqslant \alpha\boldsymbol{\xi}$ for some $\alpha > 0$, and since $M = M_1 M_2 \geqslant A$, $R \geqslant 0$, we have

$$0 \leqslant M^{-1}R\mathbf{e} = M^{-1}(M - A)\mathbf{e} = \mathbf{e} - M^{-1}A\mathbf{e},$$

or, since $M^{-1} = M_2^{-1}M_1^{-1} > 0$ by Propositions 1 and 2, $0 \leqslant M_2^{-1}M_1^{-1}R\mathbf{e} \leqslant e - \alpha M_2^{-1}M_1^{-1}\xi < \mathbf{e}$. Taking $\|\ \|$ to be the norm calculated by the maximum row sum, it follows that $0 \leqslant \rho(M^{-1}R) \leqslant \|M^{-1}R\| < 1$. ■

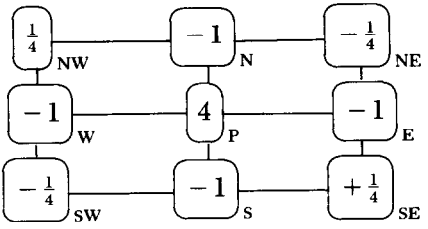
3. CONCLUDING REMARKS

It is easy to check in the following example, if the boundary conditions of (1) are replaced by Dirichlet boundary conditions, that the resulting approximation is not monotone. In this context monotonicity is equivalent to the discrete maximum principle (Ciarlet [5]).

EXAMPLE (This is due to Bramble and Hubbard [2]). $\Omega = [0,1]\times[0,1]$, $h = \frac{1}{2}$, $a = 1$, $b = 1$, $c = 1$, and $d = 0$, so that

$$Lu := -(u_{xx} + u_{xy} + u_{yy}). \tag{5}$$

h^2L^h is given by the following stencil:



Define W_{ij} by $W(\mathrm{P}) = 1$, $W(\mathrm{NW}) = -4$, $W(Q) = 0$ for other Q's. Then $L^hW_{ij} = 0$, but W has a positive interior maximum.

Bramble and Hubbard [2] have shown that if the boundary conditions for this special operator L in (5) are homogeneous Dirichlet and the boundary nodes are then eliminated from the resulting linear system, then the resulting interior submatrix of A is monotone. We conjecture that this is true more generally, but have been so far unable to extend the factorization of Proposition 2 up to the nodes adjacent to the boundary.

When L^h is replaced by L^h_θ, $0 \leqslant \theta \leqslant 1$, the proof of Theorem 1 is virtually unchanged. Only the R.H.S. of the equations (4) and (their corresponding solution) must be slightly modified.

W. Layton thanks Dr. Bem Cayco for a stimulating discussion on an earlier version of this paper and Professor R. Varga for calling his attention to the theory of monotone and oscillation matrices.

REFERENCES

1 J. H. Bramble, On the convergence of difference approximations for second order, uniformly elliptic operators, in *Numerical Solution of Field Problems in Continuum Physics* (Proc. Symp. Appl. Math., Duram, N.C., 1968), SIAM-AMS Proc., Vol. II, Amer. Math. Soc. Providence, R.I., 1970, pp. 201–209.

2 J. H. Bramble and B. E. Hubbard, New monotone type approximations for elliptic problems, *Math. Comp.* 18:349–367 (1964).

3 J. H. Bramble and B. E. Hubbard, On a finite difference analogue of an elliptical boundary value problem which is neither diagonally dominant nor of non-negative type, *J. Math. Phys.* 43:117–132 (1964).

4 A. Brandt, Generalized local maximum principles for finite-difference operators, *Math. Comp.* 27:685–718 (1973).

5 P. G. Ciarlet, Discrete maximum principle for finite difference operators, *Aequationes Math.* 4:338–352 (1970).

6 L. Collatz, Bemerkungen zur Fehlerabschatzung für das Differenzenverfahren bei Partieller Differentialgeichungen, Z. *Angew. Math. Mech.* 13:56–57 (1933).

7 G. Forsythe and W. Wasow, *Finite-Difference Methods for Partial Differential Equations*, Wiley, New York, 1960.

8 F. P. Gantmacher and M. G. Krein, Oscillation Matrices and Small Vibrations of Mechanical Systems, GITTL, Moscow, 1950; English Transl.: Office of Tech. Service, Dept. of Commerce, Washington, D.C.

9 S. Gershgorin, Fehlerabschatzung für das Differenzenverfahren zur Losung Partieller Differentialgleichungen, Z. *Angew. Math. Phys.* 10:373–382 (1930).

10 D. Greenspan and P. C. Jain, On the Nonnegative Difference Analogues of Elliptic Differential Equations, M.R.C.-T.S.R. 490, Madison, 1964.

11 A. R. Mitchell and D. F. Griffiths, *The Finite Difference Method in Partial Differential Equations*, Wiley, New York, 1980.

12 H. S. Price, Monotone and oscillation matrices applied to finite difference approximations, *Math. Comp.* 22:489–4516 (1968).

13 R. S. Varga, *Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, N.J., 1962.

Received 2 June 1986; revised 4 February 1987